DOCUMENT RESUME

ED 229 404 TM 830 300

AUTHOR Barzansky, Barbara; And Others
TITLE Applying the Concept of Trustworthiness to the Evaluation of a Clinical Program.
PUB DATE Apr 83
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Clinical Experience; *Credibility; *Evaluation Criteria; Gynecology; Higher Education; Interviews; Medical Schools; Medical Students; Observation; Obstetrics; Physicians; *Program Evaluation; *Student Teacher Relationship; *Tutorial Programs

ABSTRACT

An attending tutor program designed to increase faculty-student contact within an Obstetrics and Gynecology clerkship was evaluated. Sessions were observed, written documents were reviewed, and faculty and students were interviewed in order to determine if the program was meeting its goals. Based on the evaluation data, the program was restructured to increase the tutors' responsibility for student evaluation. The evaluation design was analyzed utilizing the criteria of trustworthiness proposed by Guba. Certain issues which may have relevance for investigators using naturalistic approaches in clinical settings were raised. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

PRINTED IN THE U.S.A.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED229404

APPLYING THE CONCEPT OF TRUSTWORTHINESS TO THE EVALUATION OF A CLINICAL PROGRAM

Barbara Barzansky
Marjorie Alschuler
C.R.B. Beckmann
Eta Berner
Burleigh Angle
Greg Ozog

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
B. Barzansky
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Center for Educational Development and
University of Illinois College of Medicine
University of Illinois at Chicago

Paper presented at the Annual Meeting of the
American Educational Research Association, April, 1983

TM 830 300

An important element common to the training of all health professionals is the application of knowlege, skills, and attitudes in a clinical practice setting. Evaluators interested in this phase of training often raise questions that are most appropriately addressed from the perspective of naturalistic inquiry (Guba and Lincoln, 1981). This paradigm assumes that there are multiple realities inherent in situations and that single variables cannot be isolated for study but must be considered as part of a complex whole (Guba, 1981). In medical education, investigators working within this paradigm often attempt to describe the experiences of students in clinical clerkships (Schermerhorn, et al., 1980) or during residency training (Bosk, 1979).

Evaluators accepting the science-based assumptions of the rationalistic paradigm (Guba, 1981) have had explicit methodological guidelines to follow in planning and determining the worth of their investigations for some time (Campbell and Stanley, 1963). In contrast, evaluators choosing the epistemological framework of naturalistic inquiry have been without ground-rules until recently. Now, the criteria developed as part of the concept of trustworthiness (Guba, 1981) can be utilized to both plan and judge evaluations conducted within the naturalistic inquiry paradigm.

The four characteristics that must be satisfied for a naturalistic inquiry to be considered trustworthy are credibility (which corresponds to internal validity in rationalistic terminology), transferability (which corresponds to generalizability), dependability (which corresponds to realiabilty), and confirmability (which corresponds to objectivity). There are threats to the achievement of each of these which must be countered by specific methodological steps. The presence of these steps in the evaluation design is the basis for judging the results of any evaluation

to be trustworthy. The minimum strategies required to ensure trustworthiness are "triangulation and member checks (for credibility), thick description (for transferability), leaving an audit trail (for dependability), and triangulation and practicing reflexivity (for objectivity)" (Guba, 1981, p.98). "Triangulation" means using multiple data sources to confirm each item of information, "member checks" involve getting feedback from the group supplying the data, "thick description" requires developing a detailed picture of the program being evaluated, leaving an "audit trail" involves collecting the documentation and a record of the process so they can be examined by an external reviewer, and "practicing reflexivity" requires each evaluator to reveal the assumptions underlying the questions asked and the conclusions reached.

This paper: 1) describes the evaluation of an educational program developed to increase faculty-student contact within a clinical clerkship, 2) applies the criteria of trustworthiness to the evaluation process, and 3) identifies some characteristics of health professions settings that may impact on an evaluator's ability to apply the criteria as completely as theoretically desirable. The ultimate aim is to begin to develop ways to circumvent these identified problem areas to ensure the maximum rigor possible in this type of evaluation.

Methodology

The Program

During the Summer of 1981, a department of Obstetrics and Gynecology initiated an "Attending Tutor" program as part of their eight-week clerkship. This was done in response to student requests for more contact with faculty. Groups of three or four students were assigned to pairs of

faculty with instructions to meet for at least one hour per week. The objectives, format, and content of the sessions were left up to the individual tutors. The first three times the tutor program was offered, the only data collected were from a student questionnaire distributed at the end of the clerkship. This assessed how satisfied students were with the program by asking them to rate it on a 1 to 5 scale. During the fourth offering of the program, a team of consultants collaborated with the faculty member responsible for planning the program to implement a systematic evaluation.

Design of the Study

The consultants and the faculty planner decided that all tutor sessions during an eight-week clerkship should be observed. Each of the five consultants chose one of the five groups of students to follow throughout the rotation. Prior to the start of the program, a meeting was held to introduce the consultants to the attending tutors. At this session, the tutors were told that their sessions would be observed and any concerns about this were discussed. Faculty were assured that what would result would be a description of the program, not an evaluation of individuals as teachers. Eight of the 10 tutors were present at the meeting and all of these agreed to participate.

In actuality, 30 of the 41 sessions that took place during the clerkship were observed. In two cases, the tutors met with their three or four students as a pair. In the other three cases, tutors met with students separately. Five to 10 sessions were observed for each group of students. Sessions lasted from 45 minutes to 1½ hours.

The consultants determined that there should be no initial constraints on the observations. The only guidelines were to record, in as much detail

as possible, what occurred during the sessions and to prepare fieldnotes for sharing as soon as possible (usually within 48 hours) after an observation. The team agreed to meet weekly.

During the first three weeks of the program, the consultant debriefing sessions concentrated on the format and content of fieldnotes. A colleague of the consultants, familiar with observational methodologies, attended most of these sessions to raise questions and make suggestions. Extensive minutes were taken by an observer at these meetings. These discussions led to more focused observations to ensure consistency in the data collected so that groups could be compared.

By the middle of the program, a set of issues/concerns was generated (e.g., what did the students and faculty think was the goal of the program, were things discussed in the sessions that were not covered elsewhere in the clerkship). These issues/concerns formed the basis of an interview guide for use with faculty and students. Some of the same questions were asked of faculty and students so that the perceptions of a tutor and his/her students could be compared.

Interviews (lasting from 30 minutes to one hour) were held with each group of students the last week of the program. Also during that week, students filled out the regular clerkship evaluation form and another questionnaire that asked them to react to the content covered during the tutor sessions. This instrument was developed by combining the topics that each observer had recorded. Students were asked whether a topic was covered in their group, whether they would have liked it discussed, and whether it was addressed elsewhere in the clerkship.

Nine of the ten tutors were interviewed at the end of the program. Two of these individuals had not been observed during the program at their request. Also at this time, any documents pertaining to the tutor program

were collected.

In summary, the study design consisted of a series of alternating data collection and data processing steps: (1) making initial observations, (2) sharing fieldnotes and initial impressions, (3) conducting focused observations, (4) generating issues/concerns, (5) developing interview guides and (6) holding interviews.

Data Analysis

Plans for data analysis were developed by the consultants in collaboration with the faculty planner. An outside observer, experienced in group process, was present at these sessions to make suggestions and to provide feedback to the group. Another peer, who was not directly involved in the study, also provided input during data collection and processing.

The results of the study were reported to the department by a two-stage process. First, a seminar highlighting the findings was presented. The questions, reactions, and suggestions generated at this presentation were incorporated into a final, detailed written report which was sent to all department members.

In organizing the study results, an attempt was made to use multiple sources of data (written documents, observations, interviews, and questionnaires) to substantiate each finding. Whenever possible, quantitative data as well as descriptive examples were used to document interpretations. As noted, the combined use of both qualitative and quantitative methods of data collection provided richer information than each could have generated alone.

Results

The results are organized into two parts. First, some major findings

that emerged from the investigation and the impact of the study on the tutor program format are summarized. Next, the results of the "evaluation of the evaluation" (i.e., the application of the criteria of trustworthiness to the evaluation design) are presented.

The Program

There were several purposes for the tutor program mentioned in documentary material. Along with enhancing contact between students and faculty to allow them to "get to know each other better" in an informal setting, the sessions should allow faculty to act as role models by demonstrating professional behavior and good clinical thinking. Also, the sessions were described as a place where faculty could evaluate student progress on a continuous basis, identify students in academic difficulty and initiate remediation, and supervise student acquisition of the clerkship objectives. When interviewed, all of the faculty but one felt that the purpose of the sessions was to get to know the students in a comfortable, non-threatening environment. Several also mentioned the role modeling function. Goals relating to evaluation and remedial education were not mentioned. When asked, all but two attendings felt that students should not be evaluated on their particpation in the sessions.

As might be expected from the flexible nature of the program and the faculty and students' belief that the purpose was to enhance interaction, there were considerable differences among the groups in the topics discussed, the teaching style of the tutors, and whether the groups met on a regular basis or sporadically. Although there was variability among groups, the observations documented that there was consistency over time within each group with respect to the format of the sessions, the degree of dominance of the attendings, and whether the attendings or students

initiated the discussion topics. While it was theoretically possible for each tutor to adopt a varied teaching style (e.g., problem-solving sessions, demonstrations, lectures), this did not occur. During the departmental seminar, faculty confirmed that they were consistent from clerkship to clerkship as well as within a clerkship.

A total of 71 discussion topics were recorded in the combined notes of all observers. These topics were classified into four categories: affective/ethical issues (for example personal reactions to death, abortion) (11 topics), medicine as a career (for example specialty choice, academic medicine vs. private practice) (19 topics), educational issues (10 topics), and specific clinical cases (31 topics). Many of these topics were covered in more than one group. Of the 12 topics that students reported were addressed in the tutor sessions and nowhere else in the clerkship, six came from the medicine as a career category and six came from educational issues. Eleven of these unique topics were discussed in at least three of the five tutor groups. The final unique topic was a clinical case. Students reported satisfaction with all of the topics covered and considered them to be useful. However, students were often inaccurate in recalling which topics were discussed within their groups. During the interview, some faculty suggested that there should be increased faculty-student contact within the clinical setting, where actual rather than hypothetical cases could be discussed.

Approximately six months after the completion of the evaluation, the structure of the tutor program was altered. Faculty were given two new and explict responsibilities: (1) supervise and evaluate the students in their groups as they perform at least one complete history and physical on a patient, and (2) conduct problem-solving sessions around clinical cases. The

tutor's role in the final clerkship evaluation was also clarified and expanded. However, the tutors were still encouraged to pursue those topics that were identified as the unique contribution of the program.

Applying the Criteria of Trustworthiness

Table 1 summarizes the methods employed in the study design to ensure the trustworthiness of the findings. Each criterion mentioned as being essential has been addressed in some degree.

TABLE 1 - APPLYING THE CRITERIA OF TRUSTWORTHINESS TO THE STUDY METHODOLOGY

CHARACTERISTIC	CRITERION	METHODOLOGY
CREDIBILITY	TRIANGULATION	-Multiple data sources (faculty, students, records) -Different methods (interviews, observation, questionnaires)
	MEMBER CHECKS	-Consultations with faculty planner -Presentation of results to faculty
TRANSFERABILITY	THICK DESCRIPTION	-Detailed fieldnotes -Probing interviews -Findings synthesized in final report
DEPENDABILITY	AUDIT TRAIL	-Peer observer present at meetings -Detailed minutes kept of meetings -Peer comment on fieldnotes and final report
CONFIRMABILITY	TRIANGULATION	-Multiple data sources -Different methods
	REFLEXIVITY	-Observers shared assumptions and interpretations

Discussion and Conclusions

During the attempt to apply the criteria of trustworthiness to this evaluation, certain issues emerged which may have relevance for other investigators using naturalistic approaches in clinical education settings. These

are more specific than the usual concerns about entry and confidentiality, which also have relevance here. Some of the issues are described below, grouped under the criteria upon which they have bearing.

Member checks: The format of the presentation to the department was a formal lecture, as this was considered most appropriate by the faculty planner. Data concerning such things as the number of topics and content covered in a session or the formats used were reacted to by those present. However, findings or interpretations that touched on professional values/ identity or academic autonomy were avoided. In these areas, the consultants could not obtain feedback on their perceptions. The question arises as to whether the formal nature of the member check was the inhibiting factor or whether the reluctance to touch on aspects of the professional role goes deeper.

Thick description: Only one clerkship rotation was observed. Faculty stated that their groups varied greatly from offering to offering depending on student characteristics such as the presence of a strong student leader in the group, excessively quiet or excessively verbal groups, or level of student experience in clinical settings. Also, this program was part of a larger clerkship experience. Topics often came from cases that students had seen in the course of their clinical responsibilities. Students also had contact with tutors in other settings (in the clinics, on the wards) which colored their behavior in and preceptions of the tutor program. This means that a given program may be impacted upon by elements outside itself and that evaluators need to understand these influences in order to fully "describe" it.

Practicing reflexivity: Some of the subjects relating to medical practice and the medical profession covered in the tutor sessions were

emotionally involving for the observers. Analyzing how those reactions influenced the observations was an important aspect of data processing but was difficult to do.

An important, perhaps self-evident, issue that arises from the use of Guba's criteria is the resource implications this has for a project. In studies of medical education, the use of observational methodologies require a considerable time commitment since the hours students spend are long and the experiences they have are varied (Klintberg and Givner, 1977). Added to this, there is now an imperative that the evaluators and one or more peers not directly associated with the project engage in frequent debriefings and keep extensive records both of content (e.g., field-notes, documents) and of process.

In summary, this paper has used the evaluation of part of a medical clerkship as a vehicle to begin to raise some issues about the process of natrualistic inquiry in clinical settings. While the evaluation may be considered a success both from the standpoints of process (i.e., most of Guba's criteria were addressed in the methodology) and of outcome (i.e., the data were utilized by the deparment to implement changes that brought the program more into line with the original goals), the basic issues raised were not fully resolved. An awareness of these issues may help in the design of future evaluations.

References

Bosk, Charles. Forgive and Remember The University of Chicago Press, Chicago, 1979.

Guba, Egon. Criteria for Assessing the Trustworthiness of Naturalistic inquiries. ECTJ 29: 75-91, 1981.

Guba, Egon and Lincoln, Y. Effective Evaluation. Jossey-Bass Publishers, San Francisco, 1981.

Klintberg, I. and Givner, N. The Comparability of Clerkship Experiences in a Variety of Locations. J. of Med. Educ 52: 518-520, 1977

Schermerhorn, G., Williams, R., Miller, R., Kienzler, L., and Wilson, S. Programme Evaluation: An Anthropological Look at the Clerkship. Medical Teacher 2: 171-174, 1980.

TM 830 300

ABSTRACT

Applying the Concept of Trustworthiness to the Evaluation of a Clinical Program

Barbara Barzansky, Ph.D., Marjorie Alschuler, Ph.D., C.R.B. Beckmann, M.D., Eta Berner, Ed.D., Burleigh Angle, M.A. and Greg Ozog, Ph.D.

Center for Educational Development, University of Illinois at Chicago

An attending tutor program designed to increase faculty-student contact within an Obstetrics and Gynecology clerkship was evaluated. Sessions were observed, written documents were reviewed, and faculty and students were interviewed in order to determine if the program was meeting its goals. Based on the evaluation data, the program was restructured to increase the tutors' responsibility for student evaluation.

The evaluation design was analyzed utilizing the criteria of trustworthiness proposed by Guba. Certain issues which may have relevance for investigators using naturalistic approaches in clinical settings were raised.